The Proportional Hazard Model: Estimation and Testing using Price Change and Labor Market Data

Fernando Alvarez
University of Chicago

Katarína Borovičková
New York University

Robert Shimer
University of Chicago

February 15, 2015
Preliminary

Abstract

We use labor market data and data on price changes to examine the role of structural duration dependence and heterogeneity in shaping the aggregate hazard rates. In line with an extensive literature we examine this question through the lens of a mixed proportional hazard model. While we think that this model is a convenient representation of the data, we recognize that its structure can be too restrictive. We focus on environments where we observe two observations per individual as this not only allows us to estimate the model non-parametrically, but also test whether the true data-generating process is likely to have a structure imposed by a mixed proportional hazard model. We reject that this is the case both for the price change data and labor market data. We then turn to data simulated from reasonable structural models, none of which can be represented as a mixed proportional hazard model, to examine implications of estimating a misspecified mixed proportional hazard model. We use a "CalvoPlus" model for price changes, while for the labor market data, we assume that individual durations follow an inverse Gaussian distribution. We find that, in fact, the mixed proportional hazard model is a good approximation of the CalvoPlus model and therefore the estimated baseline hazard rate is very similar to the true structural hazard rate. This is not the case for the inverse Gaussian model for the labor market where the mixed proportional hazard model cannot be viewed as a good approximation. As a consequence, fitting a mixed proportional hazard model to these data vastly understate the importance of heterogeneity in the economy.

1 Introduction

The longer a worker has been out of work, the less likely he is to return to work in the near future. The longer he has been working, the less likely he is to lose his job. The more time has passed since a firm has last changed its price, the less likely the firm is to change its price in the near future. These facts are well-known, as is the difficulty in interpreting them: even if every individual finds a job, loses a job, or changes a price at a constant rate, heterogeneity across individuals can give rise to the patterns described here. The composition of the "surviving" population, individuals who have not experienced such an event, changes endogenously as time passes.

This paper uses large labor market and price data sets to reexamine the role of structural duration dependence and ex ante heterogeneity in shaping aggregate hazard rates. Following much of the literature on duration models, we examine the data through the lens of the proportional hazard model. The model specifies that the hazard rate of exiting a state (finding a job, losing a job, or changing a price) is the product of two functions: a function of an individual's observed and unobserved characteristics; and a function of the duration in the state. We view this as a convenient statistical representation of the data but recognize that the multiplicative structure is restrictive and potentially incorrect.[1]

In much of our analysis, we focus on environments in which we observe two spells for each individual. We also assume that each individual's characteristics are constant across the two spells. We discuss the plausibility and restrictiveness of these assumptions in the body of the paper. Honoré (1993) proves that the proportional hazard model is nonparametrically identified under these assumptions. In particular, consider a data set containing the duration of two completed spells for each individual and no other covariates. Assume that the probability that an individual's spell ends in period t conditional on not having ended prior that period can be expressed as $\theta h(t)$, where θ is a function of the individual's characteristics and $h(t)$ is the common *baseline hazard rate*. Normalizing the population mean value of θ to unity, we can recover both the distribution of θ, say $G(\theta)$, and the entire baseline hazard $\{h(t)\}$ from this data set.

We use two large data sets, one containing price data from the United States and the other containing employment data from Austria, to estimate a proportional hazard model. We start by following the literature and imposing parametric restrictions on the distribution of individual characteristics G. In contrast to the robustness issues highlighted in Heckman and Singer (1984b) but consistent with the findings in Nakamura and Steinsson (2008), our

[1] See (see Van den Berg, 2001, Section 4.3) for an attempt to write down an economic model that generates non-constant but proportional hazard rates as an outcome.

results are robust to a variety of parametric restrictions on G.

We then offer the first nonparametric estimates of G. More precisely, we prove that the model is overidentified: if the proportional hazard assumption is valid, there are many equivalent nonparameteric estimates of baseline hazard rate. Unfortunately, we find that different estimates give us different results. That is, we can easily reject the hypothesis that our data sets could have been generated by any proportional hazard model in which each individual's characteristic is constant across the two spells.

Almost any model will be rejected in a sufficiently large data set, even if the model offers a close approximation to reality. We therefore turn to quantitatively-reasonable synthetic data in order to evaluate the effect of imposing the proportional hazard assumption in environments where it is invalid. To do this, we develop structural models that we believe may be reasonable approximations to the data generating process for both the price and labor market models. For the price data, we look at a "CalvoPlus" model (Nakamura and Steinsson, 2010), while for the labor market data, we assume individual durations follow an inverse Gaussian distribution, as in Alvarez, Borovičková, and Shimer (2014). Neither structural model has a proportional hazard representation. Instead, we suggest a multiplicative decomposition of the evolution of the raw hazard rate into the portion attributable to the average change in individuals' hazard rates (the structural hazard rate) and the proportion attributable to changes in the composition of individuals in the population. This decomposition can be performed in any structural model, including the proportional hazard model. Whenever the proportional hazard assumption is valid, the structural hazard rate is equivalent to the baseline hazard rate.

Since we know the true model, we can compute the structural hazard rate. We are interested in what would happen if we didn't know the true model. To that end, we create large synthetic data sets using the structural model and estimate the proportional hazard model both parametrically and nonparametrically. We compare the estimated baseline hazard rate with true structural hazard rate in order to evaluate the economic relevance of the proportional hazard assumption.

In the case of price data, we find that the parametrically estimated baseline hazard and the structural hazard rate are very similar. This is not the case for the labor market data, where the baseline hazard assumption induces us to vastly understate the role of heterogeneity. Our intuition is that this is related to the structural models we use in the two cases. The CalvoPlus model we use for price data implies that all hazard rates are upward sloping and quickly asymptote to a maximum value which differs across goods. Although the proportional hazard assumption is inaccurate, it is a reasonable approximation to our calibrated model. The inverse Gaussian model we use for labor market data implies that hazard rates are hump-shaped and may have peaks at very different durations. Fitting a proportional hazard to data generated from hazards that are far-from-proportional causes us to vastly understate the importance of heterogeneity in the economy.

We also consider another approach to nonparametric identification in the proportional hazard model, using covariates as in Elbers and Ridder (1982) and Heckman and Singer (1984a). More precisely, assume that an individual's hazard is the product of three terms, a baseline hazard, an unknown function of observed covariates, and an individual fixed effect which is orthogonal to the covariates. Then Elbers and Ridder (1982) and Heckman and Singer (1984a) prove that, under certain regularity conditions, the model is nonparametrically identified. As in the previous case, we prove that if the model is correctly specified, there are many equivalent nonparametric estimates of the model. Our empirical results are also broadly similar: we reject the proportional hazard assumption and find that the economic relevance of the rejection depends on how far the true model is from the proportional hazard model.

Our paper proceeds as follows...

2 Identification and Testing with Two Spells

2.1 Continuous Time

This section proves that the proportional hazard model is overidentified with data on two spells. Our approach is based on Honoré (1993), who establishes that the model is nonparametrically identified.

We consider a population with measure 1. Each individual has a fixed type θ with cumulative distribution $G(\theta)$ in the population. Each individual experiences two completed spells. We assume that the probability that a spell ends prior to period t is

$$F(t;\theta) \equiv 1 - e^{-\theta \int_0^t h(\tau) d\tau}$$

for some nonnegative, integrable function h. These outcomes are independent across spells and across individuals. Equivalent, $\theta h(t) \equiv F_t(t;\theta)/(1 - F(t;\theta))$ is the instantaneous hazard of a job ending during period t and $e^{-\theta \int_s^t h(\tau) d\tau}$ is the probability of a job ending between any dates s and t conditional on the job surviving until at least date s.

As written, this model is not identified. We could double θ for all individuals and halve $h(\tau)$ at all durations without changing the probability distribution over any individual's realized duration. We address this through a convenient normalization, that the population mean of θ is unity, $\int \theta dG(\theta) = 1$.[2] This implies that the baseline hazard at duration 0 is simply equal to the population hazard rate at that duration. At later durations, however, dynamic selection reduces the value of θ and so affects the evolution of the population hazard rate.

A key object for this analysis is the survivor function. Let $\Phi(t_1, t_2)$ denote the fraction of individuals whose first spell lasts at least t_1 periods and second spell lasts at least t_2 periods. The structure of the proportional hazard model implies that this is

$$\Phi(t_1, t_2) = \int e^{-\theta(Z(t_1)+Z(t_2))} dG(\theta), \tag{1}$$

where $Z(t) \equiv \int_0^t h(\tau)d\tau$ is the *integrated* baseline hazard. This formula takes advantage of the fact that the durations of the two spells are independent conditional on the individual characteristic θ.

Honoré (1993) proves that the model is nonparametrically identified. Denote the partial derivative of Φ using subscripts. Simple algebra implies

$$\Phi_1(t_1, t_2) = -h(t_1) \int \theta e^{-\theta(Z(t_1)+Z(t_2))} dG(\theta),$$
$$\Phi_2(t_1, t_2) = -h(t_2) \int \theta e^{-\theta(Z(t_1)+Z(t_2))} dG(\theta).$$

In particular, taking ratios of these two numbers gives

$$\frac{\Phi_1(t_1, t_2)}{\Phi_2(t_1, t_2)} = \frac{h(t_1)}{h(t_2)} \tag{2}$$

for all t_1 and t_2. Thus the survivor function contains enough information to recover the ratio of the baseline hazard rate at any two durations. In particular, the baseline hazard rate at duration 0 is equal the population hazard rate, say $h(0) = \bar{h}(0)$, while the baseline hazard at later durations satisfies

$$h(t) = \bar{h}(0) \frac{\Phi_1(t, 0)}{\Phi_2(t, 0)}.$$

Effectively this approach treats the distribution of unobserved characteristics as a nuisance parameter and solves for the baseline hazard by differencing out the nuisance parameter.

Once we have recovered the baseline hazard $h(t)$, we immediately obtain the integrated baseline hazard $Z(t)$. We can then recover the Laplace transformation of the distribution of individual characteristics θ. This follows immediately from the survivor function in equa-

[2]If the distribution of θ does not have a finite mean, this assumption is not a normalization. Nevertheless, the results in Proposition 1 still holds at any strictly positive values of t_1, t'_1, and t_2.

tion (1). For any number s,

$$\mathcal{L}(s) = \int e^{-\theta s} dG(\theta) = \Phi(Z^{-1}(s), 0),$$

where the left hand side is the Laplace transform of G and the right hand side is given by data and the inverse of the integrated baseline hazard.[3]

We take Honoré's argument one step further. Evaluate the ratio of partial derivatives at two values (t_1, t_2) and (t_1', t_2). Taking ratios again gives

$$\frac{h(t_1)}{h(t_1')} = \frac{\Phi_1(t_1, t_2)\Phi_2(t_1', t_2)}{\Phi_2(t_1, t_2)\Phi_1(t_1', t_2)} \tag{3}$$

for all t_1, t_1', and t_2. Curiously, the left hand side does not depend on t_2, while the right hand side depends on t_2, a testable prediction of the proportional hazard model:

Proposition 1 *For any t_1 and t_1',*

$$\Psi(t_1, t_1'; t_2) \equiv \frac{\Phi_1(t_1, t_2)\Phi_2(t_1', t_2)}{\Phi_2(t_1, t_2)\Phi_1(t_1', t_2)}$$

does not depend on t_2.

In fact, Honoré (1993) considers a more general model in which the baseline hazard rate is allowed to differ across the two spells. In that case, equation (2) gives the ratio of the baseline hazard during the first spell at duration t_1 relative to the baseline hazard in the second spell at duration t_2, while we need equation (3) to recover the relative value of the baseline hazard at two durations during the same spell.

Proposition 1 yields a nonparametric test of the model. $\Psi(t_1, t_1'; t_2)$ can be measured directly in a large dataset for a particular value of t_1 and t_1', and different values of t_2. The proportional hazard model implies that it should be independent of t_2. This result is intuitive. In general, the relative hazard at durations t_1 and t_1' during the first spell depends on individual's characteristics and hence is correlated with the duration of the second spell. But this is not the case with the proportional hazard model, since everyone has the same relative hazard at durations t_1 and t_1'.

[3]This is feasible as long as the integrated baseline hazard grows without bound; otherwise we can only do this for small values of s.

2.2 Discrete Time

The results in Proposition 1 assume that we have data available in continuous time, while any real world data set records outcomes at discrete intervals. There are two ways to deal with this. The first is to assume that the time period is sufficiently short so that we can move freely between continuous and discrete time. The second is to develop a discrete time analog of this result. We pursue the latter route here.

Assume time is discrete, $t \in \{1, 2, \ldots\}$. Let $\theta h(t)$ denote the probability that an individual with type θ finds a job during period t. The probability that a spell ends prior to t is

$$F(t;\theta) = 1 - \prod_{\tau=1}^{t-1}(1 - \theta h(\tau)),$$

independent across spells and individuals. Again let $\Phi(t_1, t_2)$ denote the fraction of individuals whose first spell lasts at least t_1 periods and second spell lasts at least t_2 periods. This now satisfies

$$\Phi(t_1, t_2) = \int \left(\prod_{\tau=1}^{t_1-1}(1 - \theta h(\tau))\right)\left(\prod_{\tau=1}^{t_2-1}(1 - \theta h(\tau))\right) dG(\theta).$$

First differencing this object yields

$$\Phi_1(t_1, t_2) = -h(t_1)\int \theta \left(\prod_{\tau=1}^{t_1-1}(1 - \theta h(\tau))\right)\left(\prod_{\tau=1}^{t_2-1}(1 - \theta h(\tau))\right) dG(\theta),$$
$$\Phi_2(t_1, t_2) = -h(t_2)\int \theta \left(\prod_{\tau=1}^{t_1-1}(1 - \theta h(\tau))\right)\left(\prod_{\tau=1}^{t_2-1}(1 - \theta h(\tau))\right) dG(\theta),$$

where now $\Phi_1(t_1, t_2) \equiv \Phi(t_1 + 1, t_2) - \Phi(t_1, t_2)$ and $\Phi_2(t_1, t_2) \equiv \Phi(t_1, t_2 + 1) - \Phi(t_1, t_2)$. We can again eliminate the nuisance parameter to arrive at equations (2) and (3). Thus once we have properly defined the proportional hazard model, the testable implications of the discrete and continuous time versions of this model are identical.

2.3 Formal Test Statistics

In any real-world data set generated from a proportional hazard model, we would not expect $\Psi(t_1, t_1'; t_2)$ to be exactly independent of t_2. We plan to develop a formal test statistic for the null hypothesis that this is true.

3 General Decomposition of the Hazard Rate

The proportional hazard model offers a natural decomposition of the raw hazard rate into the baseline hazard and the portion attributable to heterogeneity. This section shows how to perform a more general decomposition of the raw hazard rate into two portions, the average change in individuals' hazard rates (the structural hazard rate) and the remaining portion attributable to changes in the composition of individuals in the population.

Consider a population composed of many individuals characterized by a characteristic θ. Let $h(t;\theta)$ denote the hazard rate of type θ and duration t. In the proportional hazard model, this can be expressed as the product of θ and the baseline hazard, but we relax that restriction here. The probability that a spell lasts at least t periods is

$$F(t;\theta) = 1 - e^{-\int_0^t h(\tau;\theta)d\tau}$$

for all t and θ. This implies the density of θ in the population surviving to duration t is

$$g(\theta;t) = \frac{e^{-\int_0^t h(\tau;\theta)d\tau} g(\theta;0)}{\int e^{-\int_0^t h(\tau;\theta')d\tau} g(\theta';0)d\theta'},$$

where $g(\theta;0)$ is the initial population density of θ.[4] This model then gives rise to a raw hazard rate

$$H(t) = \int h(t;\theta)g(\theta;t)d\theta. \tag{4}$$

We are interested in decomposing the evolution of H into two terms, the portion due to the evolution of the average value of h and the portion due to the evolution of g. We propose a multiplicative decomposition here, $H(t) = H^{str}(t)H^{het}(t)$, or equivalently $\log H(t) = \log H^{str}(t) + \log H^{het}(t)$.[5]

We start by differentiating equation (4):

$$\dot{H}(t) = \int \dot{h}(t;\theta)g(\theta;t)d\theta + \int h(t;\theta)\dot{g}(\theta,t)d\theta.$$

The first term (if negative) is the decrease in the raw hazard rate coming from the fact that the average individual has an decreasing hazard rate. The second term (if negative) is the decrease in the hazard rate coming from the fact that individuals with a high hazard rate become a small share of the population over time. To perform a multiplicative decomposition,

[4] With a general formulation of the hazard rate h, there is no loss of generality in assuming that the initial distribution admits a density function.

[5] See Alvarez, Borovičková, and Shimer (2014) for an additive decomposition. We use a multiplicative decomposition here because it has the same structure as the proportional hazard model.

we divide through by both sides by the hazard rate and write

$$\frac{\dot{H}(t)}{H(t)} = \frac{\dot{H}^{str}(t)}{H^{str}(t)} + \frac{\dot{H}^{het}(t)}{H^{het}(t)},$$

where

$$\frac{\dot{H}^{str}(t)}{H^{str}(t)} = \frac{\int \dot{h}(t;\theta)g(\theta;t)d\theta}{H(t)} \text{ and } \frac{\dot{H}^{het}(t)}{H^{het}(t)} = \frac{\int h(t;\theta)\dot{g}(\theta;t)d\theta}{H(t)}.$$

Finally, we define

$$\log H^{str}(t) = \int_{t_0}^{t} \frac{\dot{H}^{str}(s)}{H^{str}(s)} ds + \log H(t_0) \text{ and } \log H^{het}(t) = \int_{t_0}^{t} \frac{\dot{H}^{het}(s)}{H^{str}(s)} ds$$

for some carefully chosen value of t_0, e.g. $t_0 = 0$.

In the proportional hazard model, $h(t;\theta) = \theta h(t)$, so

$$\frac{\dot{H}^{str}(t)}{H^{str}(t)} = \frac{\dot{h}(t)}{h(t)}.$$

Thus this decomposition recovers the baseline hazard rate, $H^{str}(t) = h(t)$. More generally, however, the growth in the structural portion of the hazard rate represents the increase in the average hazard rate relative to the average level of the hazard rate. In a structural model, we can compute the contribution of structural duration dependence $H^{str}(t)$ and compare it to what one would get by treating the data as if it comes from a proportional hazard model.

4 Data

4.1 Sampling Framework

To estimate the proportional hazard model, we need data on the survivor function $\Phi(t_1, t_2)$ for a large variety of "individuals." An issue that immediately arises is that in any real-world data set, we do not observe two completed spells for all individuals. For example, in a labor market context, there are individuals who never work and others who only stop working when they hit retirement.

Our methodology recognizes that if the proportional hazard model is correctly specified, then we can estimate it using any subset of observations in the data. For example, in the context of price changes, we can estimate it only using retailers who stock a good for at least a pre-specified amount of time. While this may bias any estimates of the distribution of characteristics $G(\theta)$, it should not affect estimates of the baseline hazard $h(t)$. Our methodology also takes advantage of the symmetry of our setup. This is important because we may not observe the second spell for individuals whose first spell is very long, but we can observe the first spell of individuals whose second spell is very long.

Our goal is to measure the baseline hazard through to some pre-specified duration T. For the case of price changes, our sampling frame is the set of all products that are in our data set for at least $2T-1$ periods after the initial price change. For each product, we let t_1 equal the duration of the first spell, top-coded at T. For each product with $t_1 < T$, we let t_2 equal the duration of the second spell, again top-coded at T. This is feasible because we have at least $2T-1$ observations and because we do not look at products whose first spell is top-coded. Denote the the number of products with durations (t_1, t_2) by $n(t_1, t_2)$ for all $t_1 < T$ and $t_2 \leq T$. Let $n(T, \cdot)$ denote the number of products whose first spell has duration at least T.

For $t_1 < T$ and $t_2 < T$, our measure of the number of spells is simply $N(t_1, t_2) = (n(t_1, t_2) + n(t_2, t_1))/2$, where we take advantage of the symmetry of our model to effectively enlarge the data set. For $t < T$, we also let $N(t, T) = N(T, t) = n(t, T)$, again using symmetry, but now to impute values for individuals whose first spell is top-coded and hence second spell may be truncated. Finally, our measure of $N(T, T)$ is $n(T, \cdot) - \sum_{t<T} n(t, T)$, i.e. the remaining spells.

Once we have recovered $N(t_1, t_2)$ for all $(t_1, t_2) \in 1, 2, ..., T^2$, we can define the survivor function as

$$\Phi(t_1, t_2) = \frac{\sum_{\tau_1 \geq t_1, \tau_2 \geq t_2} N(\tau_1, \tau_2)}{\sum_{\tau_1 \geq 1, \tau_2 \geq 1} N(\tau_1, \tau_2)}, \tag{5}$$

the fraction of individuals with spells lasting at least (t_1, t_2) periods. This is an unbiased estimate of the survivor function for the original set of products and hence we can use it to estimate the baseline hazard, recover the distribution of characteristics, and test the model.

Our sampling framework for the labor market application is slightly different. We still measure two consecutive spells, top coding each spell at duration T. But spells are no longer consecutive, since a worker spends some time employed between unemployment spells and vice versa.[6] We therefore restrict attention to individuals whom we observe for at least $4T$ periods, still top-coding both spells at T and inferring the duration of the second spell for individuals whose first spell is top-coded from the individuals whose second spell is top-coded.

[6] A similar consideration applies if we consider regular price changes, rather than all price changes, sine sales can occur in between regular price changes.

4.2 Price Data

We use the Nielsen-IRI retail scanner data sets, which are available through the Kilts Center for Marketing at the University of Chicago. This contains a large number of weekly price observations for many products in many retail outlets. The products are divided into 31 categories, e.g. coffee and frozen entrees. We estimate and test the proportional hazard on each variety.

4.3 Labor Market Data

For our labor market application, we use data from the Austrian social security registry. The data set covers the universe of private sector workers over the years 1972–2007 (Zweimuller, Winter-Ebmer, Lalive, Kuhn, Wuellrich, Ruf, and Buchi, 2009). It contains information on individual's employment, registered unemployment, maternity and retirement, with the exact begin and end date of each spell.

The use of the Austrian data is compelling for two reasons. First, the data set contains the complete labor market histories of the majority of workers over a 35 year period, which allows us to construct multiple non-employment spells per individual. Second, the labor market in Austria remains flexible despite institutional regulations, and responds only very mildly to the business cycle. Therefore, we can treat the Austrian labor market as a stationary environment and use the pooled data for our analysis. We discuss the key regulations below.

Almost all private sector jobs are covered by collective agreements between unions and employer associations at the region and industry level. The agreements typically determine the minimum wage and wage increases on the job, and do not directly restrict the hiring or firing decisions of employers. The main firing restriction is the severance payment, with size and eligibility determined by law. A worker becomes eligible for the severance pay after three years of tenure if he does not quit voluntarily. The pay starts at two month salary and increases gradually with tenure.

The unemployment insurance system in Austria is very similar to the one in the U.S. The duration of the unemployment benefits depends on the previous work history and age. If a worker has been employed for more than a year during two years before the layoff, she is eligible for 20 weeks of the unemployment benefits. The duration of benefits increases to 30 weeks and 39 weeks for workers with longer work history.

Temporary separations and recalls are prevalent in Austria. Around 40 percent of non-employment spells end with an individual returning to the previous employer.

We work with non-employment spells, defined as the time from the end of one full-time job to the start of the following full-time job during which a worker was registered as unemployed for at least one day. We drop incomplete spells and spells involving a maternity leave. Although in principle we could measure non-employment duration in days, disproportionately many jobs start on Mondays and end on Fridays, and so we focus on weekly data. We code 0 to 6 days out of work and 0 weeks, 7 to 13 days as 1 week, and so on.

Our sample consists of all individuals who were no older than 45 in 1972 and no younger than 40 in 2007. Thus each individual has at least 15 years when he could potentially be at work. For each individual, we take the first two completed non-employment spells satisfying the following criterion: the age at the start of the spell must be at least 25, the duration of the spell must be between 0 and 260 weeks. We drop all individuals who do not have two such spells. In this notation, the duration of 0 weeks means that the spell started and ended in the same calendar week. We drop spells longer than 260 weeks (5 years) so that we have a consistent sample across all years. Workers with few years in the data set are unlikely to have very long spells.

We also work with employment spells, defined as the time between non-employment spells. For this purpose, we also consider a worker to be non-employed if he spends at least xxx weeks out of full time employment.

5 Parametric Estimates

6 Non-Paremetric Test Results

We implement the non-parametric test for the case with two spells using both data on price changes as well as non-employment exit rates (employment exit rate to be added).

6.1 Price Changes

We start with the price data. In Figure 1 we show results for only one good category, coffee, but the results are similar (often even stronger) for other goods. Each line in the figure shows the probability of finding a job at duration t_1 relative to the hazard at duration 2 weeks, estimated using different lengths of the second spell t_2.

Figure 2 shows a subset of the results, the relative probability of finding a job at durations 13, 26, 39, and 52 weeks, compared to 2 weeks. According to the theory, these probabilities should not depend on the choice of t_2, and so should give accurate estimates of the relative baseline hazard $h(t_1)/h(2)$, but the figure shows a systematic dependence. Each line initially increases and then starts declining at some $t_2 < t_1$. The maximum implied relative baseline hazard is in each case at least twice the minimum. Monte Carlo simulations suggest to us

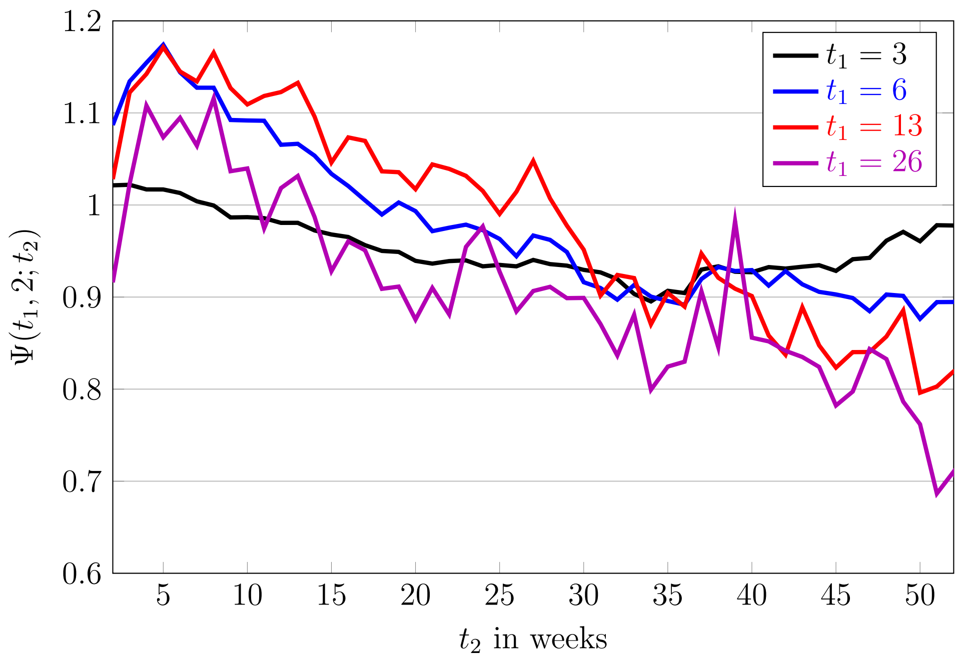

Figure 1: Test of the proportional hazard model for the price data, good category coffee. The figure shows the baseline probability of changing prices at 3, 6, 13, and 26 weeks, compared to 2 weeks duration for different values of the second spell duration t_2. According to the model, each line should be independent of t_2.

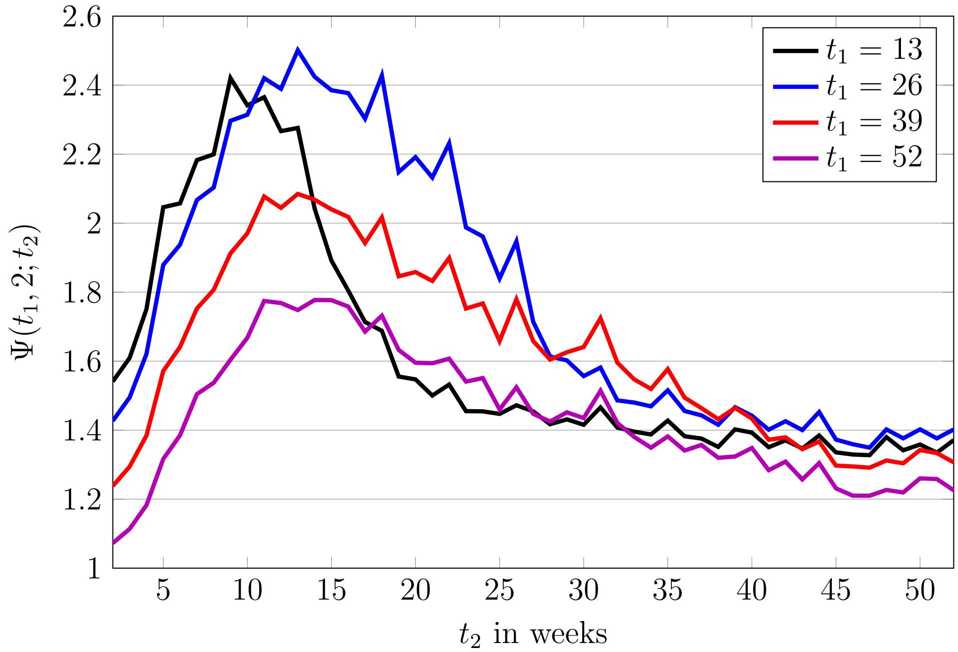

Figure 2: Test of the proportional hazard model. The figure shows the baseline probability of finding a job at 13, 26, 39, and 52 weeks, compared to 2 weeks duration for different values of the second spell duration t_2. According to the model, each line should be independent of t_2.

that this is driven by the large number of individuals who have two spells of similar long lengths, an observation that cannot be accommodated by the proportional hazard model.

7 Structural Models

The test results show that the neither the price data, nor the labor market data are likely to be generated by the proportional hazard model. But what does it mean to fail the test? What goes wrong if one ignores the test results and estimates the proportional hazard model? Using structural models we argue that this leads to incorrect inference about the role of the structural dependence versus heterogeneity.

7.1 CalvoPlus Model of Price Changes

We model the decision of a firm to change the product price using the so-called CalvoPlus model which is a combination of the menu cost and Calvo model. A firm can adjust its price at any time but has to pay the fixed cost $\psi > 0$. However, occasionally, the firm can change its price for free. This opportunity arrives at the rate $\lambda > 0$.

Let $P(t)$ be the current price and $p^*(t)$ the profit maximizing price, which follows a Brownian motion with zero drift and variance σ^2. A firm then optimally adjusts its price either when it is free (Calvo model), or when the gap between the current optimal price is large enough (menu costs model). More precisely, a firm adjusts its price when the price gap $x(t) \equiv p^*(t) - P(t)$ hits either $-w/2$ or $w/2$, where w is the width of the inaction region which depends on ψ, σ. A firm always adjust to the profit-maximizing price.

We can now turn to determination of the spell without a price adjustment. All spells start at the profit-maximizing price, therefore $x(0) = 0$. The price gap $x(t)$ then follows a Brownian motion with zero drift and variance σ^2, properties inherited from $p^*(t)$. The spell without price adjustment then ends either because an opportunity of free price adjustment arrives, or $x(t)$ hits one of the barriers. In the second case, the length of a price spell is given by the first hitting time of either of the two barriers for Brownian motion, formula for which is known (see for example Kolkiewicz (2002)), and is depends only on $\theta \equiv (\sigma/w)^2$.

The distribution of times between two price adjustments depends only on θ and λ. In particular, the probability of not adjusting for at least t periods is a product of two terms - probability that Calvo does not arrive before time t and probability that $x(t)$ does not hit any of the two barriers,

$$S(t; \theta) = e^{-\lambda t} \sum_{j=0}^{\infty} \frac{a_j}{q_j} e^{-t\theta q_j},$$

where $q_j = (2j+1)^2\pi^2/2$ and $a_j = 2\pi(-1)^j(2j+1)$ for $j = 0, 1, \ldots$.

We consider economy with many products, each described ψ, σ, λ. We impose that λ is the same for each product and allow ψ, σ to vary across products, denoting the distribution of θ as G. The aggregate survivor function can be obtained by integrating individual $S(t; \theta)$ across all types,

$$S(t) = e^{-\lambda t} \sum_{j=0}^{\infty} \frac{a_j}{q_j} \int e^{-t\theta q_j} dG(\theta) = e^{-\lambda t} \sum_{j=0}^{\infty} \frac{a_j}{q_j} \mathcal{L}_G(tq_j)$$

where $\mathcal{L}_G$ is the Laplace transform of distribution G.

It turns out to be particularly convenient to assume that $\theta \sim Gamma(k, \nu)$, with Laplace transform $\mathcal{L}_G(z) = (1 + \nu z)^{-k}$. We show in Appendix that a very good approximation for the aggregate hazard $H(t)$ is given a simple formula

$$H(t) \approx \frac{\pi^2/2k\nu}{1 + \pi^2/2\nu t}.$$

Observe that this is not a proportional hazard model because $h(0, \theta) = \lambda$ for each θ

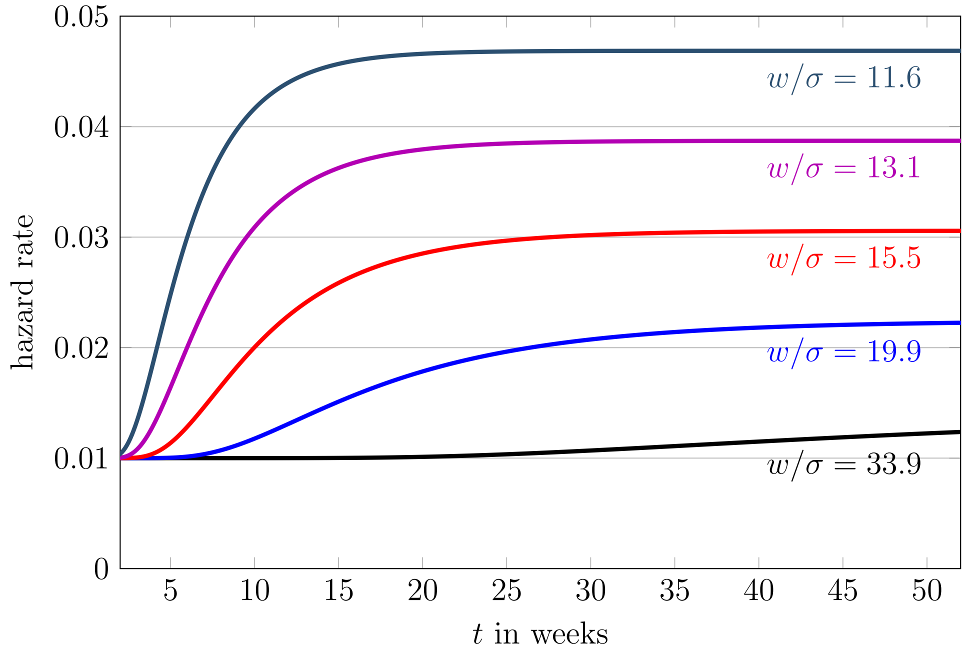

Figure 3: Type-specific hazard rates for CalvoPlus model. The figure shows hazard rate of a price change for $\lambda = 0.01$ and different values of normalized width of inaction w/σ.

but $h(t, \theta) \neq h(t', \theta)$ for $t \neq t', t > 0, t' > 0$. In words, the hazard rate of adjusting price at duration $t = 0$ is given by the hazard rate of having the opportunity of free price adjustment, which is the same for each product. After duration $t = 0$, the hazard rate of price adjustment is determined by the probability that the price gap hits the barrier, which in depends on θ. This is illustrated in Figure 3, which shows the hazard rates of price adjustment for $\lambda = 0.01$ and different values of θ.

We choose parameters of the model λ, ν, k to match the aggregate hazard rate for one of the product, coffee. The comparison of the hazard rate implied by the model and that of the data is depicted in Figure 4. We find that $\lambda = 0.01, \nu = 0.01, k = 2.25$ which implies that the mean value of w/σ is 8.11, with standard deviation of 3.87.

We use the parametrized model to analyze the following exercise. We use data generated by the CalvoPlus model and use them to estimate the proportional hazard model, despite the fact the model is misspecified. We then evaluate how the proportional hazard model performs by comparing the proportional hazard-implied contribution of the structural duration dependence to the aggregate hazard rate (i.e. the baseline hazard) to the true one. We compute the true contribution of heterogeneity using our multiplicative decomposition formula XXX. We can do this because we know the true structure of the model. Figure 5 shows that, despite the fact that the model is misspecified, the baseline hazard rate coincides

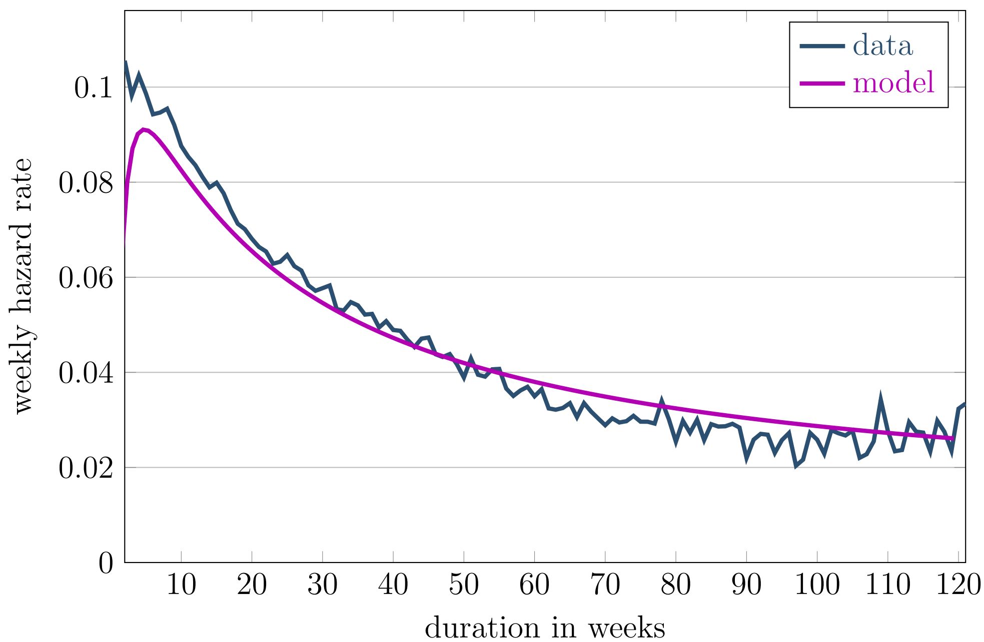

Figure 4: Fit of the CalvoPlus model to the data for coffee. The figure compares the aggregate hazard rate of price adjustment implied by the model to the one measured in the data.

with the true contribution of the structural duration dependence.

To interpret this result, we come back to our non-parametric test. We perform the test on the data simulated by the model, and show $\Psi(t_1, 9; t_2)$ if Figure 6 for different values of t_1. We again reject that the data are generated by the proportional hazard model, but the rejections is not as strong.

Indeed, going back to Figure 3, we notice that the pattern in the individual hazard rates. They all start at λ for $t = 0$, then they are increasing until they reach their type-specific asymptote. Once on the asymptote, the hazard rate is constant and thus this is a version of the proportional hazard model with constant hazard and non-degenerate distribution of heterogeneity. Most types reach their asymptote very quickly, within 10 weeks or so, and thus the model is a proportional hazard model at longer durations.

This exercise also illustrates that the non-parametric test is "too strong." The proportional hazard model is rejected, as shown in 7, despite the fact that the model is rather close to the proportional hazard model. A different way to see that the model is rejected is to look at the non-parametric estimates of the baseline hazard rate. If the proportional hazard model is correct, then $\hat{\bar{h}}(t|t_2) = S_1(t, t_2)/S_2(t, t_2)$ is a consistent estimate (up to a scale) of the baseline hazard rate for any choice of t_2. Figure 7 shows $\hat{\bar{h}}(t|t_2)$ for different values of t_2, normalized such that they equal aggregate hazard rate at $\bar{t} = 10$.

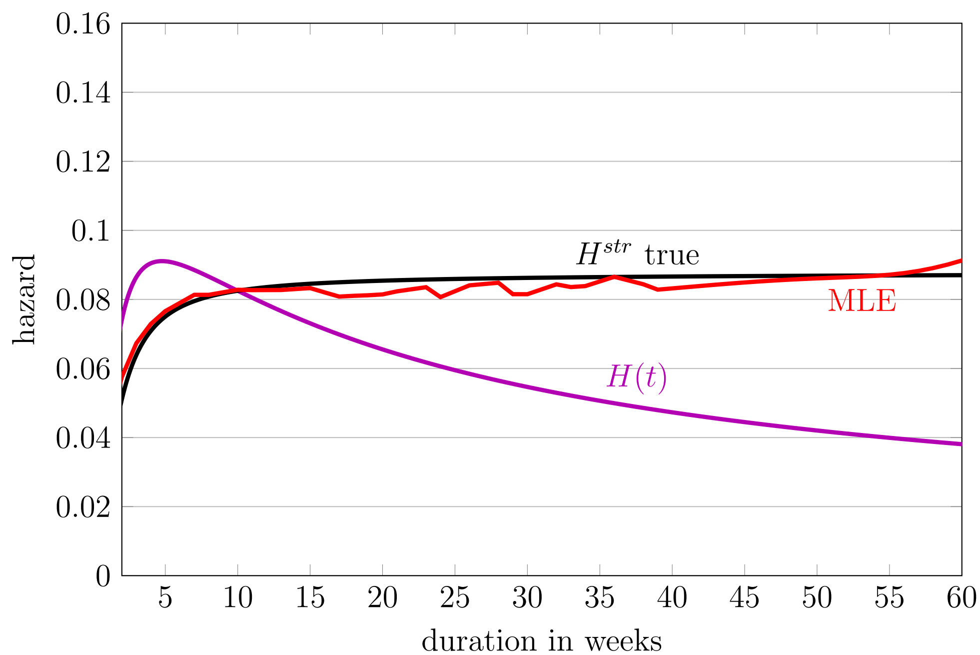

Figure 5: Multiplicative hazard rate decomposition in the CalvoPlus model. The purple line shows the aggregate hazard rate. The black line shows the true contribution of structural duration dependence, H^{str}, to the hazard rate, while the red line shows the contribution of the structural duration dependence implied by proportional hazard model, estimated using random effects and a parametric baseline hazard. Both are normalized such that at $\bar{t} = 10$ they are equal to the aggregate hazard rate.

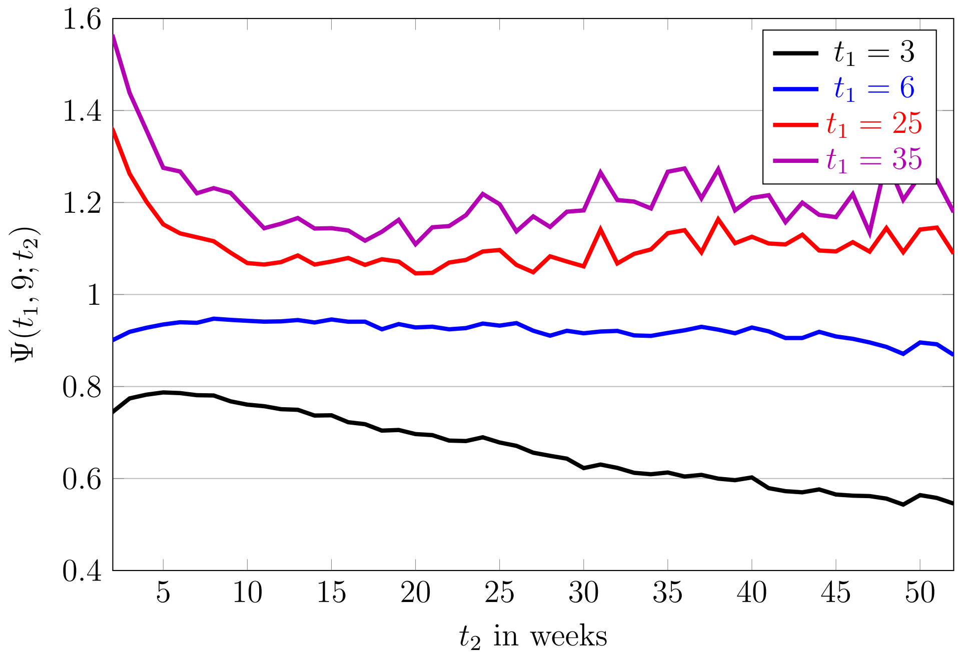

Figure 6: The non-parametric test of the proportional hazard model using data simulated from the CalvoPlus model.

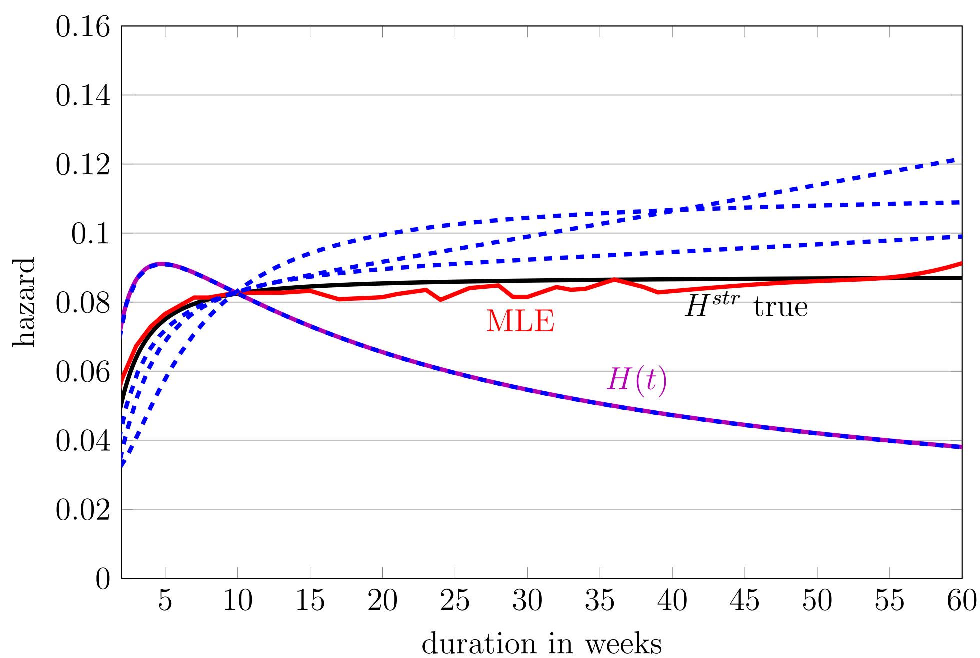

Figure 7: The non-parametric test of the proportional hazard model using data simulated from the CalvoPlus model.

7.2 Inverse-Gaussian Model of Non-employment Exit

We perform a similar analysis for the non-employment exit. We use a structural model for the decision to exit non-employment to generate artificial data which we use to estimate the proportional hazard model. We use the structural model from Alvarez, Borovičková, and Shimer (2014).

Consider a risk-neutral, infinitely-lived worker with discount rate r, who can either be employed, $s(t) = e$, or non-employed, $s(t) = n$, at each instant in continuous time t. We assume the worker earns a flow wage $e^{w(t)}$ when working and gets a flow benefit $e^{b(t)}$ when not working. The natural logarithm of the potential wage $w(t)$ and of the potential benefit $b(t)$ follow correlated random walks with drift, both when the worker is employed and when the worker is non-employed. The drift and standard deviation of each may depend on the worker's employment status. We impose restrictions on these parameters to ensure that the worker's value is finite.

A non-employed worker can become employed at t by paying a fixed cost $\psi_e e^{b(t)}$ for a constant $\psi_e \geq 0$. Likewise, an employed worker can become non-employed by paying a cost $\psi_n e^{b(t)}$ for a constant $\psi_n \geq 0$. The worker must decide optimally when to change her employment status $s(t)$.

Let $\omega(t) \equiv w(t) - b(t)$ denote the worker's log net benefit from employment. This inherits the properties of w and b, following a random walk with state dependent drift and volatility. We prove in Alvarez, Borovičková, and Shimer (2014) that the worker's employment decision depends only on her employment status $s(t)$ and her net benefit from working $\omega(t)$. In particular, the worker's optimal policy involves a pair of thresholds. If $s(t) = e$ and $\omega(t) \geq \underline{\omega}$, the worker remains employed, while she stops working the first time $\omega(t) < \underline{\omega}$. If $s(t) = n$ and $\omega(t) \leq \bar{\omega}$, the worker remains non-employed, while she takes a job the first time $\omega(t) > \bar{\omega}$. Assuming the sum of the fixed costs $\psi_e + \psi_n$ is strictly positive, the thresholds satisfy $\bar{\omega} > \underline{\omega}$, while the thresholds are equal if both fixed costs are zero.

We turn next to the determination of non-employment duration. All non-employment spells start when an employed worker's wage hits the lower threshold $\underline{\omega}$. The net benefit from employment then follows a Brownian motion with drift and the non-employment spell ends when the net benefit from employment hits the upper threshold $\bar{\omega}$. Therefore the length of a non-employment spell is given by the first passage time of a Brownian motion with drift. This random variable has an inverse Gaussian distribution with density function

$$f(t; \alpha, \beta) = \frac{\beta}{\sqrt{2\pi} t^{3/2}} \exp\left(-\frac{(\alpha t - \beta)^2}{2t}\right), \tag{6}$$

where the parameters α and β depend on the structural parameters of the model. In par-

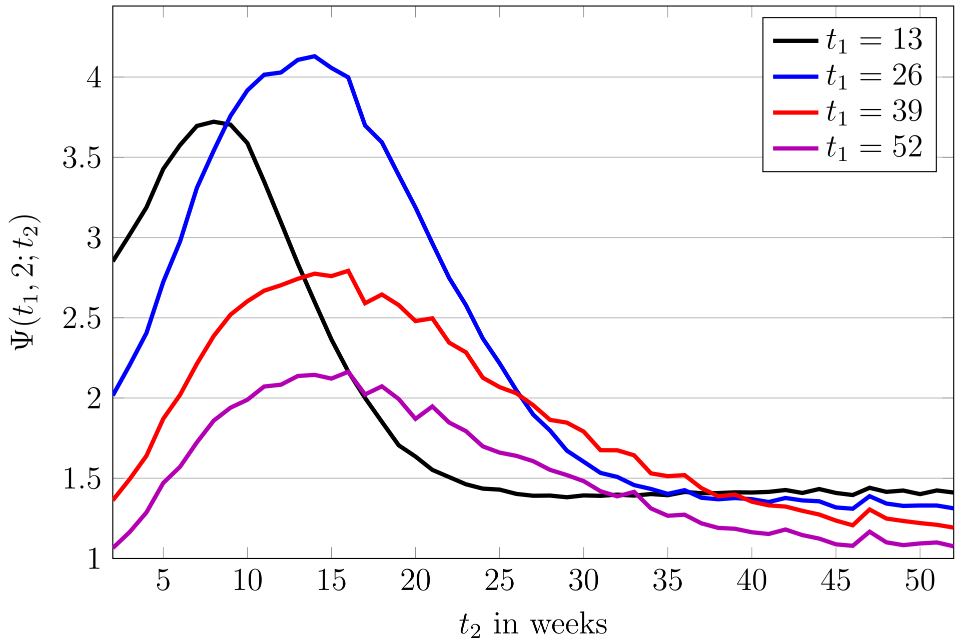

Figure 8: The non-parametric test of the proportional hazard model using data simulated from Inverse-Gaussian model.

ticular, α is the ratio of the drift of the Brownian motion to its standard deviation and β is the ratio of the distance between the barriers to the standard deviation.

We show in Alvarez, Borovičková, and Shimer (2014) that using data on two non-employment spells, we can non-parametrically estimate the distribution of $(|\alpha|, \beta)$, call it G. Here we take the estimate distribution G and simulate data from this model. We again non-parametrically test the proportional hazard model. Figure 8 shows a strong rejection of the model, with patterns very similar to those observed in 2.

The rejection for the proportional hazard model is not surprising, here the non-employment exit hazard rates differ a lot across types. Each hazard rate is hump-shaped, starting at 0 at duration 0, and approaching asymptotic hazard rate of $\alpha^2/2$ for $t \to \infty$. The position and the height of the maximum of hte hazard rate, as well as the speed at which it approaches its asymptote depend on values of α, β.

Despite the strong rejection of the proportional hazard model, we estimate baseline hazard and compare it to the true contribution of the structural duration dependence $H^{str}(t)$, which is calculated using the structure of the model. These are compared in Figure 10. The baseline hazard misses the true structural duration dependence by a wide margin.

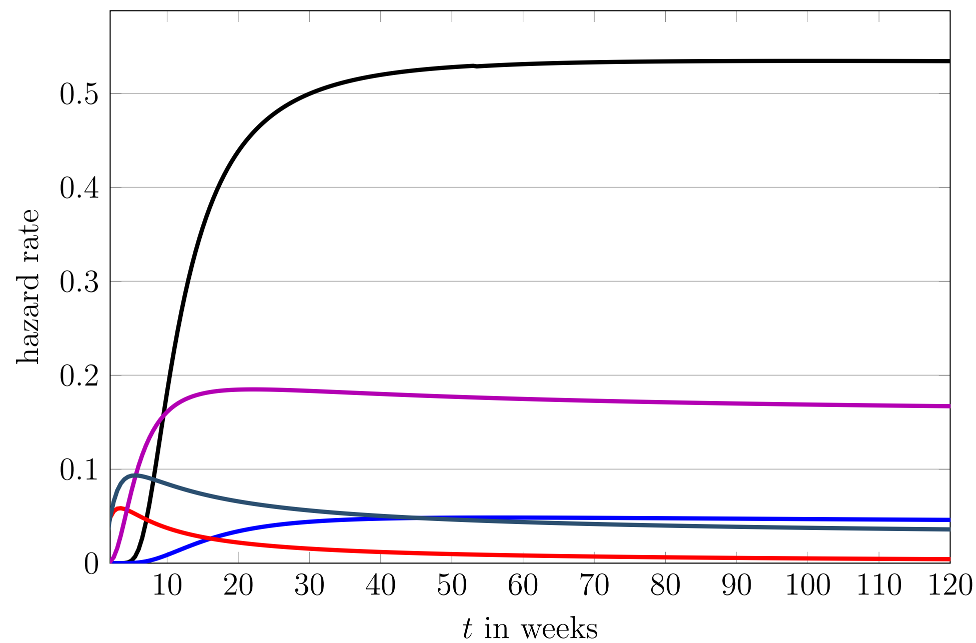

Figure 9: Type-specific hazard rates for the Inverse Gaussian model. The figure shows hazard rate of exiting non-employment for different types described by (α, β).

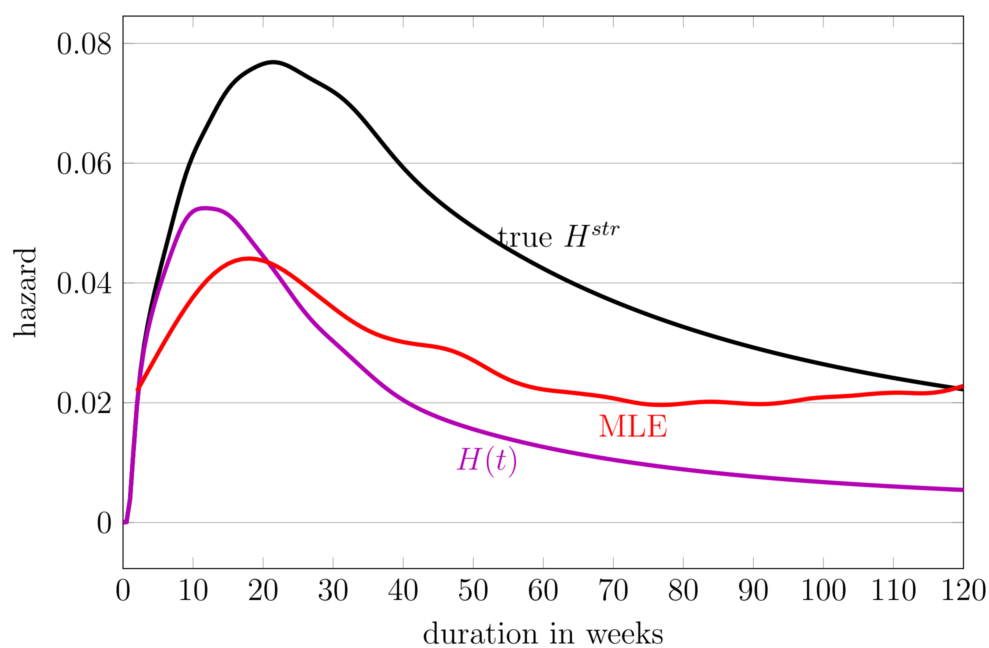

Figure 10: Multiplicative hazard rate decomposition in the Inverse Gaussian model. The purple line shows the aggregate hazard rate. The black line shows the true contribution of structural duration dependence, H^{str}, to the hazard rate, while the red line shows the contribution of the structural duration dependence implied by estimated proportional hazard model. Both are normalized such that at $\bar{t} = 2$ they are equal to the aggregate hazard rate.

8 Covariates

Elbers and Ridder (1982) show that covariates help to identify the proportional hazard model. In particular, they show that if the mean of G is finite, and the function ϕ is non-negative, differentiable and non-constant on an open set in R^k, where k is the number of covariates, then the functions ϕ and $\bar{h}$ and the distribution G are uniquely determined up to a constant. Heckman and Singer (1984a) give another identifiability theorem, which relaxes the assumption of the finite mean, but instead require a condition on the rate of decay of the tail of G.

We first show how the survivor function can be used to determine $\phi(x)$ and $\bar{h}(x)$, and how it can be used to derive a test of the model. Our approach here is closer to Elbers and Ridder (1982). In fact, after some algebra it can be shown that our results for $\phi(x)$ and $\bar{h}(x)$ are equivalent to equations (10) and (15) in Elbers and Ridder (1982).

This section is written in a continuous time and continuous space, as is also the case in Elbers and Ridder (1982) and Heckman and Singer (1984a). We believe that it is not possible to conduct the same analysis in when time or space are discrete, as it is not possible to eliminate the distribution of θ.

Assume now that the hazard of an individual i is given by

$$h_i(t) = \theta_i \psi(x_i) \bar{h}(t) \tag{7}$$

where x_i is an observable characteristic of individual i, as for example age.

Let $S(t,x)$ be the share of individuals with characteristic x for whom the spell lasts at least t periods. Then,

$$S(t,x) = \int \exp\left(-\theta\psi(x)\int_0^t \bar{h}(s)ds\right) g(\theta)d\theta. \tag{8}$$

Differentiate with respect to t,

$$S_t(t,x) = -\psi(x)\bar{h}(t)\int \theta \exp\left(-\theta\psi(x)\int_0^t \bar{h}(s)ds\right) g(\theta)d\theta. \tag{9}$$

Evaluate this expression at $t = 0$, $S_t(0,x) = -\psi(x)\bar{h}(0)\int \theta g(\theta)d\theta$. As usually, the baseline hazard is identified up to a scale and thus we can normalize $\bar{h}(0) = \int \theta g(\theta)d\theta = 1$. Equation (9) then identifies $\psi(x)$. This result is analogous to equation (10) in Elbers and Ridder (1982).

Differentiate equation (8) with respect to x to find

$$S_x(t,x) = -\psi'(x)\int_0^t \bar{h}(s)ds \int \theta \exp\left(-\theta\psi(x)\int_0^t \bar{h}(s)ds\right) g(\theta)d\theta. \tag{10}$$

Take the ratio of $S_t(t,x)$ and $S_x(t,x)$,

$$\frac{S_t(t,x)}{S_x(t,x)} = \frac{\psi(x)\bar{h}(t)}{\psi'(x)\int_0^t \bar{h}(s)ds}. \tag{11}$$

Define $y(t) \equiv \int_0^t \bar{h}(s)ds$, and observe that equation (11) is an ordinary differential equation for $y(t)$, with the initial condition $y(0) = \int_0^0 \bar{h}(s)ds = 0$. Rewriting (11) in terms of $y(t)$,

$$y'(t) = y(t)\frac{\psi'(x)}{\psi(x)}\frac{S_t(t,x)}{S_x(t,x)} \equiv k(t,x)y(t), \tag{12}$$

we find the solution for $y(t)$,

$$y(t) = Ct\exp\left(\int_0^t \left(k(\tau,x) - \frac{1}{\tau}\right)d\tau\right), \tag{13}$$

where C is a constant to be determined. By taking a logarithmic transformation of the above expression and differentiating, it is straightforward to verify that this is indeed a solution.

The condition $y(0) = 0$ does not help to pin down the value of C. However, notice that the normalization $h(0) = 1$ implies another condition for $y(t)$, specifically that $y'(0) = h(0) = 1$. Take the derivative of equation 13 with respect to t,

$$\bar{h}(t) = y'(t) = Ctk(t,x) = C\frac{t\psi(x)S_t(t,x)}{\psi'(x)S_x(t,x)}. \tag{14}$$

The value at $t = 0$ can be found by taking a limit as $t \to 0$,

$$\bar{h}(0) = \lim_{t\to 0} C\frac{t\psi(x)S_t(t,x)}{\psi'(x)S_x(t,x)}.$$

Recall that $S_t(0,x) = -\phi(x)$, and use L'Hospital rule to find the following limit,

$$\lim_{t\to 0}\frac{S_x(t,x)}{t} = -\lim_{t\to 0}\psi'(x)\frac{\int_0^t \bar{h}(\tau)d\tau}{t} = -\psi'(x)\lim_{t\to 0}\frac{\bar{h}(0)}{1} = -\psi'(x).$$

Therefore, it follows that $C = 1$. After some algebra, it can be shown that our solution for $y(t)$ is equivalent to the solution implied by equations (11) and (15) in Elbers and Ridder (1982).

Observe that the right hand side is directly measurable in the data so this gives one way to find the baseline hazard rate $\bar{h}(t)$.

It is possible to use equation 11 for find an overidentifying test. Take a ratio of equation (11) evaluated at (t, x) and (t', x),

$$\frac{S_t(t,x)}{S_x(t,x)} \frac{S_x(t',x)}{S_t(t',x)} = \frac{\bar{h}(t) \int_0^{t'} \bar{h}(s)ds}{\bar{h}(t') \int_0^{t} \bar{h}(s)ds}. \tag{15}$$

The right hand side of (15) does not depend on x while the left-hand side does. The left-hand side is directly measurable in the data and thus for given values of t, t' one can test whether the measured ratio is independent of x.

In the case of one spell and covariates we consider a data set that for each x has spells with at least T periods. For each x we top code the spells at duration T, i.e. for any completed or uncompleted duration $t \geq T$ we record them as of lengh T. Then for each x and $t \leq T$ we record the number of spells as $n(t, x)$. The survivor function $S(x, t)$ is obtained for each x and $t \leq T$ as:

$$S(x,t) = \frac{\sum_{s \geq t} n(x,s)}{\sum_s n(x,s)}$$

9 Conclusion

References

Alvarez, Fernando, Katarína Borovičková, and Robert Shimer, 2014. "Decomposing Duration Dependence in a Stopping Time Model." University of Chicago Mimeo.

Elbers, Chris, and Geert Ridder, 1982. "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model." *The Review of Economic Studies*. 49 (3): 403–409.

Heckman, James J., and Burton Singer, 1984a. "The Identifiability of the Proportional Hazard Model." *Review of Economic Studies*. 51 (2): 231–241.

Heckman, James J, and B Singer, 1984b. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica*. 52 (2): 271–320.

Honoré, Bo E., 1993. "Identification Results for Duration Models with Multiple Spells." *Review of Economic Studies*. 60 (1): 241–246.

Kolkiewicz, Adam W, 2002. "Pricing and hedging more general double-barrier options." *Journal of Computational Finance*. 5 (3): 1–26.

Nakamura, Emi, and Jón Steinsson, 2008. "Five Facts about Prices: A Reevaluation of Menu Cost Models." *Quarterly Journal of Economics*. 123 (4): 1415–1464.

Nakamura, Emi, and Jón Steinsson, 2010. "Monetary Non-neutrality in a Multisector Menu Cost Model." *The Quarterly Journal of Economics*. 125 (3): 961–1013.

Van den Berg, Gerard J., 2001. "Duration Models: Specification, Identification and Multiple Durations." in *Handbook of Econometrics*, vol. 5, chap. 55, pp. 3381–3460. Elsevier.

Zweimuller, Josef, Rudolf Winter-Ebmer, Rafael Lalive, Andreas Kuhn, Jean-Philipe Wuellrich, Oliver Ruf, and Simon Buchi, 2009. "Austrian Social Security Database." Mimeo.